

Formulario, Simulación II

López Pérez Olin Tonatiuh

May 2023

1. Definición función de densidad:

Una función de densidad de probabilidad caracteriza el comportamiento probable de una población en tanto especifica la posibilidad relativa de que una variable aleatoria continua \mathbf{X} tome un valor cercano a x . Una variable aleatoria \mathbf{X} tiene una función de densidad f_X siendo f_X una función no negativa integrable de Lebesgue.

$$P[a \leq X \leq b] = \int_a^b f_X dx.$$

Si F_X es la función de distribución de \mathbf{X} , entonces,

$$F_X(x) = \int_{-\infty}^x f_X(u) du$$

y si f_X es continua en x

$$f_X(x) = \frac{d}{dx} F_X(x)$$

2. Distribución uniforme:

2.1. Notación:

Si \mathbf{X} es una variable aleatoria continua con distribución uniforme se escribe $X \sim U(a, b)$ o $X \sim \text{Unif}(a, b)$.

2.2. Parámetros

$a, b \in \mathbb{R}, a < b$; a es un parámetro de ubicación y $b - a$ es un parámetro de escala.

2.3. Rango

$$[a, b]$$

2.4. Media

$$\frac{a+b}{2}$$

2.5. Varianza

$$\frac{(b-a)^2}{12}$$

2.6. Función de densidad

$$f_X(x) = \frac{1}{b-a}, \text{ para } x \in [a, b]$$

2.7. Función generadora de momentos

$$\frac{e^{tb} - e^{ta}}{t(b-a)}$$

3. Distribución normal

3.1. Notación:

Si \mathbf{X} es una variable aleatoria con distribución normal se escribe $\mathbf{X} \sim \mathbf{N}(\mu, \sigma)$
o $\mathbf{X} \sim \text{Normal}(\mu, \sigma)$

3.2. Parámetros

$$\mu \in \mathbb{R}, \sigma > 0$$

3.3. Rango

$$x \in \mathbb{R}$$

3.4. Media

$$\mu$$

3.5. Varianza

$$\sigma^2$$

3.6. Función de densidad

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

3.7. Función generadora de momentos

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

4. Distribución exponencial

4.1. Notación:

Si \mathbf{X} es una variable aleatoria continua cuya distribución es normal se escribe $\mathbf{X} \sim \mathbf{E}(\lambda)$ o $\mathbf{X} \sim \mathbf{Exp}(\lambda)$

4.2. Parámetros

$$\lambda > 0$$

4.3. Rango

$$[0, \infty]$$

4.4. Media

$$\frac{1}{\lambda}$$

4.5. Varianza

$$\frac{1}{\lambda^2}$$

4.6. Función de densidad

$$f_X(x) = \lambda e^{-\lambda x}$$

4.7. Función de distribución

$$1 - e^{-\lambda x}$$

4.8. Función generadora de momentos

$$(1 - \frac{t}{\lambda})^{-1}$$

5. Distribución Weibull

5.1. Notación:

Si \mathbf{X} es una variable aleatoria continua se dice que \mathbf{X} tiene una distribución de Weibull y se denota $\mathbf{X} \sim \mathbf{Weibull}(\alpha, \sigma)$ o $\mathbf{X} \sim \mathbf{W}(\alpha, \sigma)$

5.2. Parámetros

$$\alpha, \sigma > 0$$

5.3. Rango

$$x \in (0, \infty)$$

5.4. Media

$$\frac{1}{\lambda} \Gamma(1 + \frac{1}{\alpha})$$

5.5. Varianza

$$\frac{1}{\sigma^2} [\Gamma(1 + \frac{2}{\alpha}) - \Gamma^2(1 + \frac{1}{\alpha})]$$

5.6. Función de densidad

$$\lambda \alpha (\lambda x)^{\alpha-1} e^{-(\lambda x)^\alpha}$$

5.7. Función de distribución

$$1 - e^{-(\lambda x)^\alpha}$$

5.8. Función generadora de momentos

$$\sum_{n=0}^{\infty} \frac{t^n}{n! \lambda^n} \Gamma(1 + \frac{n}{\alpha}), \alpha \leq 1$$

6. Distribución Bernoulli

6.1. Notación:

Sea \mathbf{X} es una variable aleatoria discreta que mide el número de éxitos y se realiza un único experimento con dos posibles resultados denominados éxitos o fracaso se dice que la variable aleatoria \mathbf{X} tiene una distribución de Bernoulli y se denota $\mathbf{X} \sim \text{Bernoulli}(p)$

6.2. Parámetros

$$0 < p < 1$$

6.3. Rango

$$x \in 0, 1$$

6.4. Media

$$p$$

6.5. Varianza

$$p(1 - p)$$

6.6. Función de densidad

$$p^x(1 - p)^{1-x}$$

6.7. Función de distribución

$$\begin{cases} 0 & \text{si } x < 0, \\ 1 - p & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

6.8. Función generadora de momentos

$$1 - p + pe^t$$

7. Distribución binomial

7.1. Notación:

Si \mathbf{X} es una variable aleatoria discreta se dice que \mathbf{X} tiene una distribución binomial y se denota $\mathbf{X} \sim \mathbf{B}(\mathbf{n}, \mathbf{p})$

7.2. Parámetros

$n \geq 0$ número de ensayos, $0 \leq p \leq 1$ probabilidades de éxito

7.3. Rango

$$x \in [0, 1, \dots, n]$$

7.4. Media

$$np$$

7.5. Varianza

$$np(1 - p)$$

7.6. Función de densidad

$$\binom{n}{x} p^x (1 - p)^{n-x}$$

7.7. Función de distribución

$$I_{1-p}(n - \lfloor x \rfloor, \lfloor x \rfloor + 1)$$

7.8. Función generadora de momentos

$$(1 - p + pe^t)^n$$

8. Distribución geométrica

8.1. Notación:

Si \mathbf{X} es una variable aleatoria discreta que sigue una distribución geométrica se escribe $\mathbf{X} \sim \text{Geometrica}(\mathbf{p})$

8.2. Parámetros

$$0 \leq x \leq 1$$

8.3. Rango

$$x \in [1, 2, \dots, \infty)$$

8.4. Media

$$\frac{1}{p}$$

8.5. Varianza

$$\frac{(1-p)}{p^2}$$

8.6. Función de densidad

$$p(1 - p)^{x-1}$$

8.7. Función de distribución

$$1 - (1 - p)^{x-1}$$

8.8. Función generadora de momentos

$$\frac{p}{1 - (1-p)e^t}$$

9. Distribución de Poisson

9.1. Notación:

Si \mathbf{X} es una variable aleatoria discreta se dice que \mathbf{X} tiene una distribución de Poisson y se denota $\mathbf{X} \sim \mathbf{P}(\lambda)$ o $\mathbf{X} \sim \mathbf{Poi}(\lambda)$

9.2. Parámetros

$$\lambda \in (0, \infty)$$

9.3. Rango

$$k \in [0, 1, \dots)$$

9.4. Media

$$\lambda$$

9.5. Varianza

$$\lambda$$

9.6. Función de densidad

$$\frac{e^{-\lambda} \lambda^k}{k!}$$

9.7. Función de distribución

$$\frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor!} \text{ para } k \geq 0 \text{ (donde } \Gamma(x, y) \text{ es la función gamma incompleta)}$$

9.8. Función generadora de momentos

$$\exp(\lambda(e^t - 1))$$

10. Prueba de bondad de ajuste

La metodología es la siguiente:

1-. Se colocan los n datos históricos en una tabla de frecuencias de $m = \sqrt{n}$ intervalos, se obtiene la frecuencia observada en cada intervalo (FOi), se calcula la media y la varianza de los datos.

2-. Se propone una distribución de probabilidad de acuerdo con la tabla de frecuencias obtenida en el paso 1.

3-. Con la distribución propuesta, se calcula la frecuencia esperada para cada uno de los intervalos (FEi) mediante la integración de la distribución propuesta y su posterior multiplicación por el número total de datos.

$$FEi = n \frac{1}{b-a} \int_{Lsup}^{Linf} f(x) dx$$

4-. Se calcula el estimador

$$C = \sum_{i=1}^m \frac{(FEi - FOi)^2}{FEi}$$

5-. Si $C \leq x^2$ correspondiente con $m - k - 1$ grados de libertad (k = número de parámetros estimados de la distribución) y a un nivel de confiabilidad $1 - \alpha$, entonces no se puede rechazar la hipótesis de que los datos siguen la distribución propuesta.

11. Uso de los números aleatorios.

11.1. Montecarlo de media muestral 1

1-. Se escoge aleatoriamente una semilla X_0

2-. $X_{i+1} = (ax_i + b) \bmod(m)$

3-. $U_i = X_{i+1}/m$

11.2. Montecarlo de media muestral 2

I: Cálculo de integrales. a) Para $\theta = \int_0^1 g(x)dx \dots (1)$

Observamos que si $U \sim (0, 1)$, podemos expresar (1) cómo

$$\theta = E[g(U)]$$

12. Prueba de Medias.

Verificamos que la media sea $\frac{1}{2}$

$$H_0 : \mu = \frac{1}{2}$$
$$H_2 : \mu \neq \frac{1}{2}$$

Paso 1: Calcular la media de los n números generados.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n r_i$$

Paso 2: Calcular los límites superior e inferior de aceptación.

$$l_{sx} = \frac{1}{2} + z_{\frac{\alpha}{2}} \frac{1}{\sqrt{12n}}$$
$$l_{ix} = \frac{1}{2} - z_{\frac{\alpha}{2}} \frac{1}{\sqrt{12n}}$$

Paso 3 : Si $l_{ix} \leq x \leq l_{sx}$ aceptamos que los números tienen una media estadísticamente igual a $\frac{1}{2}$ con un nivel de aceptación $1 - \alpha$

13. Prueba de Varianza.

$$\text{Varianza} = \frac{(b-a)^2}{12}$$

$$H_0 = V(x) = \frac{1}{12}$$
$$H_i = V(x) \neq \frac{1}{12}$$

Paso 1: Calcular la varianza de los n números generados.

$$V(x) = \sum_{i=1}^n \frac{(r_i - \bar{x})^2}{n-1}$$

Paso 2: Calcular los límites superior e inferior de aceptación.

$$L_{sV}(x) = \frac{x_{\frac{\alpha}{2}, n-1}^2}{12(n-1)}$$
$$L_{sV}(x) = \frac{x_{1-\frac{\alpha}{2}, n-1}^2}{12(n-1)}$$

Paso 3 : Si $L_{iV(x)} \leq V(x) \leq L_{sV(x)}$ aceptamos la hipótesis nula y los números aleatorios tienen una varianza estadísticamente igual a $\frac{1}{12}$ con un nivel de aceptación $1 - \alpha$

14. Prueba de uniformidad

$$\begin{aligned} H_0 : r_i &\sim U(0,1) \\ H_i : r_i &\text{ no son uniformes} \end{aligned}$$

La metodología es la prueba de bondad de ajuste de Chi-cuadrada

15. Prueba de Independencia.

Para Demostrar que los números generados son estadísticamente independientes entre si se propone la hipótesis

$$\begin{aligned} H_0 : r_i &\sim \text{independiente} \\ H_i : r_i &- \text{ dependientes} \end{aligned}$$

Pruebas de corridos.

Paso 1: Clasificar cada número aleatorio con respecto al anterior, de acuerdo con .

$$\begin{aligned} si : r_i &\leq r_{i-1} & r_i &= - \\ H_i : r_i &> r_{i-1} & r_i &= + \end{aligned}$$

Paso 2: Calcular el número de corridos observados h.

Una corrida se forma por un conjunto de números aleatorios consecutivos del mismo signo

Paso 3 : Calcular $E(h)$ y $V(h)$

$$\begin{aligned} E(h) &= \frac{2n-1}{3} \\ V(h) &= \frac{16n-29}{90} \end{aligned}$$

donde n es el número de datos generados.

Paso 4 : Calcular el estadístico $z = \frac{h-E(h)}{\sqrt{V(h)}}$ si es menor que el valor crítico $z_{\frac{\alpha}{2}}$ se acepta la hipótesis de independencia.

16. Prueba de corridas.

Son pruebas estadísticas no paramétricas para verificar la aleatoriedad de datos. Utiliza series de datos para decidir si son aleatorios o no.

Paso 1 Contar el número de corridas en la secuencia de datos

H_0 : La secuencia se produjo de forma aleatoria.

H_1 : La secuencia no se produjo de forma aleatoria.

Paso 2: Usar el estadístico Z como prueba:

$$Z = \frac{R - \bar{R}}{S_R}.$$

Donde R es el número de corridas observadas \bar{R} , definido como:

$$\bar{R} = \frac{2n_1n_2}{n_1+n_2} + 1.$$

Con desviación estándar:

$$S_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

n_1 y n_2 son los números de valores positivos y negativos de la serie.

Paso 3: Comparar el valor estadístico Z calculado con el $Z_{critico}$ para un nivel de confianza dado ($Z_{critico} = 1,96$ para un nivel de confianza del 95%) la hipótesis nula se rechaza, es decir, se declara que los números no son aleatorios si $|Z| > Z_{critico}$.

17. Prueba de Kolmogorov-Smirnov.

Es recomendable para conjunto i pequeños ($i \leq 30$)

Paso 1: Ordenar de menor a mayor los números del conjunto r_i .

$$r_1 \leq r_2 \leq \dots \leq r_n$$

Paso 2: Determinar los valores de D^+ , D^- , D con las siguientes ecuaciones:

$$D^+ = \max_{1 \leq i \leq n} \left(\frac{1}{n} - r_i \right)$$

$$D^- = \max_{1 \leq i \leq n} \left(r_i - \frac{i-1}{n} \right)$$

$$D = \max(D^+, D^-)$$

Paso 3 : Determinar el valor critico $D_{\alpha,n}$ de la tabla de valores críticos de kolmogorov Smirnov para un grado de confianza α y tamaño de muestra n

Paso 4 : Si $D > D_{\alpha,n}$ se concluye que el conjunto r_i no sigue una distribución uniforme, de lo contrario se acepta que sí la sigue

18. Prueba de autocorrelación.

En esta prueba, dados los datos $x_1, x_2, x_3, \dots, x_n$ se prueba la independencia de los datos, analizando los coeficientes de correlación, r_k , de r_1 de:

$x_1, x_2, x_3, \dots, x_{n-1}$ con x_2, x_3, \dots, x_n

r_2 de:

$x_1, x_2, x_3, \dots, x_{n-2}$ con x_2, x_3, \dots, x_n

...

En general

r_k de:

$x_1, x_2, x_3, \dots, x_{n-k}$ con $x_{k+1}, x_{k+2}, \dots, x_n$

Si los datos son independientes, se espera que r_k sea poco significativo. Pero más que al cálculo de r_k se usan pruebas estadísticas, como la de Ljung-Box.

19. Generación de variables aleatorias.

19.1. distribución normal

Otra forma de generar números con distribución normal es mediante la aproximación

$$x = \frac{U^{0,135} - U(1-u)^{0,135}}{0,1975}$$

Generar $U_i, i = 1, \dots, 12$

$$\text{Hacer } x = \sum_{i=1}^{12} u_i - 6$$

Salida x

Recordar que para generar la variable en distribución $F(x)$:

Generar $U \sim U(0, 1)$

Hacer $x = F^{-1}(U)$

Salida x

19.2. Teorema central del límite

Sean $\mathbf{X}_1, \dots, \mathbf{X}_n$ variables aleatorias idénticamente distribuidas con $E[X_i] = \mu$ y $Var(X_i) = \sigma^2 < \infty$ se define:

$$Z_n := \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$$

Entonces la función de distribución de \mathbf{Z}_n converge hacia la función de distribución normal estándar cuando $n \rightarrow \infty$, es decir,

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Es muy comun encontrarlo con la variable estandarizada de \mathbf{Z}_n en función de la media muestral \bar{X} , es decir

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Puesto que son equivalentes (solo se divide tanto numerador como denominador entre n). Es importante remarcar que este teorema no dice nada acerca de la distribución de la variable aleatoria \mathbf{X}_i , excepto la existencia de media y varianza.

19.3. Teorema del tamaño de muestra

El teorema del tamaño de muestra, también conocido como teorema de la precisión de la muestra, establece que para obtener una estimación de una cierta medida de interés con un error muestral máximo dado, se requiere una muestra de tamaño mínimo determinado, que depende de la varianza poblacional, el nivel

de confianza y el margen de error deseado. Formalmente, se puede enunciar de la siguiente manera:

Dado un nivel de confianza $1 - \alpha$, un error muestral máximo ϵ , y una variable aleatoria X con varianza poblacional σ^2 , entonces el tamaño mínimo de muestra n necesario para estimar la media poblacional μ con un error muestral máximo ϵ y un nivel de confianza $1 - \alpha$ está dado por:

$$n > \left(\frac{2\sigma z_{\alpha/2}}{\epsilon} \right)^2$$

donde $z_{\alpha/2}$ es el valor crítico de la distribución normal estándar que deja una probabilidad $\alpha/2$ en la cola superior.

19.4. Teorema de la suma Gaussiana

Sean \mathbf{X} e \mathbf{Y} variables aleatorias independientes con distribución normal, con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2 , respectivamente. Entonces, las variables aleatorias $\mathbf{Z}_1 = \mathbf{X} + \mathbf{Y}$ y $\mathbf{Z}_2 = \mathbf{X} - \mathbf{Y}$ siguen una distribución normal con media $\mu_{\mathbf{Z}_1} = \mu_1 + \mu_2$ y $\mu_{\mathbf{Z}_2} = \mu_1 - \mu_2$, respectivamente, y varianza $\sigma_{\mathbf{Z}_1}^2 = \sigma_1^2 + \sigma_2^2$ y $\sigma_{\mathbf{Z}_2}^2 = \sigma_1^2 + \sigma_2^2$, respectivamente. Es decir:

$$\begin{aligned}\mathbf{Z}_1 &\sim \mathbf{N}(\mu_{\mathbf{Z}_1}, \sigma_{\mathbf{Z}_1}^2) \\ \mathbf{Z}_2 &\sim \mathbf{N}(\mu_{\mathbf{Z}_2}, \sigma_{\mathbf{Z}_2}^2)\end{aligned}$$

20. Ley de los grandes números

Engloban varios teoremas que describen el comportamiento de promedio de una sucesión de variables aleatorias conforme aumenta su número de ensayos.

20.1. Ley débil

La ley débil de los grandes números establece que si $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \dots$ es una sucesión infinita de variables aleatorias independientes que tiene el mismo valor esperado μ y varianza σ^2 entonces el promedio

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

Converge en probabilidad a μ , en otras palabras para cualquier número ϵ se tiene

$$\lim_{n \rightarrow \infty} P(|\overline{X}_n - \mu| < \epsilon) = 1$$

20.2. Ley fuerte

La ley fuerte de los grandes números establece que si $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \dots$ es una sucesión infinita de variables aleatorias independientes e idénticamente distribuidas que cumplen $\mathbf{E}[\mathbf{X}_i] < \infty$ y tienen valor esperado $E[X_i] = \mu$ entonces

$$P(\lim_{n \rightarrow \infty} \overline{X}_n = \mu) = 1$$